STAT 611 Part 3
Nonstationary Models
(chapters 3, 5, 6.4, 10)

Zepu Zhang
May 4, 2011

In this course we touch upon the following approaches to non-stationarity:

1. Non-periodic trend functions (linear, quadractic, or others).

2. Periodic trend functions (e.g. sine or cosine waves).

3. Transformation.

4. Differencing.

5. Seasonal model.

Emphasis will be placed on the last two approaches. The seasonal model approach is a special type of differencing—differencing at seasonal lags.

"Differencing" and "seasonal models" are also "transformations" of Y_t. Hence the "transformation" approach in the list above must be referring to "other" transformations, in particular, a simple, analytic transform, such as $\sqrt{Y_t}$ and $\log Y_t$ for positive processes.

The "trend function" (non-periodic or periodic) approach should be used only when there is a reasonable explanation for the reason of trend. Because time series models are usually used for forecast, and any trend function is assumed to be valid whenever the model covers (hence including the future), using a trend function without a reasonable explanation is very dangerous.

Trend expressed by deterministic functions is called "deterministic trend". Trend induced by differencing or the seasonal model is called "stochastic trend".

In addition, later we'll discuss two topics that also deal with some kind of "nonstationarity": intervention analysis (change in mean), and ARCH/GARCH models (change in conditional variance).

1 Non-periodic trends

1.1 Regression for linear and quadratic trends in time

(Part of sections 3.3, 3.4)

Suppose

$$Y_t = \mu_t + X_t = \beta_0 + \beta_1 t + X_t$$

where X_t is stationary with mean 0. β_0 and β_1 can be estimated by least square.

As we have learned in linear models, the predictor here is $1, \ldots, n$ (which is time t), the response is $Y_1, \ldots, Y_n$. The least squares estimators are

$$\hat{\beta}_1 =$$
$$\hat{\beta}_0 =$$

Example: page 31.

In STAT 401, the noise in the data is iid. In time series here, the noise (X_t) is usually correlated. Therefore the properties of these estimators are not exactly like in STAT 401.

A more convenient formula for $\hat{\beta}_1$ is (3.4.7) on p. 39.

From (3.4.7) we can study the mean and variance of $\hat{\beta}_0$ and $\hat{\beta}_1$. It turns out both are unbiased.

The variance of $\hat{\beta}_1$ is given by (3.4.9), assuming $\rho_k = 0$ for $k > 1$. Note how the value of ρ_1 affects the variance.

1.2 Interpreting regression output

(Section 3.5)

Recall linear regression model in STAT 401:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

where $\epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$.

Using regular linear regression procedure (least squares) to estimate linear trends in time series is not totally suitable, because the procedure assumes iid (and possibly normal) residuals, whereas in times series

$$Y_t = \beta_0 + \beta_1 t + X_t$$

the residuals X_t are usually not independent.

But the estimate provides some baseline to look at.

Note: in the linear regression, t are predictors and Y_t are responses.

After estimating $\hat{\beta}_0$ and $\hat{\beta}_1$, we have estimated means

$$\hat{\mu}_t = \hat{\beta}_0 + \hat{\beta}_1 t$$

(compare with $\hat{Y}_i$ in STAT 401) and estimated residuals

$$\hat{X}_t = Y_t - \hat{\mu}_t$$

(compare with e_i in STAT 401)

In STAT 401 Two other things related to "goodness of fit":

- We always get an estimate of σ^2 by

$$s^2 = \frac{1}{n-p} \sum_i (y_i - \hat{y}_i)^2$$

  where p is number of β's. In time series, this is meaningful if we know (or assume) X_t has constant mean.

- Coefficient of determination,

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}}$$

2 Periodic trends

(Part of sections 3.3, 3.4)

3 Transformations for achieving stationarity

1. If $Y_t > 0$ and its spread (standard deviation) appears to be proportional to the magnitude of Y_t, that is,

$$E(Y_t) = \mu_t, \quad \sqrt{\text{var}(Y_t)} = \sigma \cdot \mu_t$$

   Then study $\log Y_t$, because

$$\log Y_t \approx \log \mu_t + \mu_t^{-1}(Y_t - \mu_t),$$

   leading to

$$E(\log Y_t) \approx \log \mu_t, \quad \text{var}(\log Y_t) \approx \mu_t^{-2} \text{var}(Y_t) = \sigma^2$$

2. If Y_t is changing roughly exponentially, that is,

$$Y_t = (c + \epsilon_t)Y_{t-1}$$

then

$$\log Y_t = \log(c+\epsilon_t)+\log Y_{t-1} \approx \log c+\frac{(c + \epsilon_t) - c}{c}+\log Y_{t-1} = \log c+\frac{\epsilon_t}{c}+\log Y_{t-1}$$

The log series has a linear time trend.

3. If Y_t appears to have stable percentage changes, that is,

$$Y_t = (1 + X_t)Y_{t-1}$$

where X_t is small, say $|X_t| < 0.2$, then follow the above,

$$\log Y_t \approx \log Y_{t-1} + \epsilon_t.$$

Hence $\nabla \log Y_t$ is the new target to study.

4 Stationarity through differencing

Section 5.1 presents several examples of converting a nonstationary process to a stationary one by differencing.

Define $\underline{\text{first difference at lag 1}}$:

$$\nabla Y_t = Y_t - Y_{t-1}$$

and $\underline{\text{centered second difference}}$:

$$\nabla^2 Y_t = (Y_{t+1} - Y_t) - (Y_t - Y_{t-1}) = Y_{t+1} - 2Y_t + Y_{t-1}$$

Example   Random walk:
$$Y_t = Y_{t-1} + e_t.$$

Then
$$\nabla Y_t = e_t$$

is stationary.

In the next few examples we assume

$$Y_t = \mu_t + X_t$$

where μ_t changes only "slowly" with time and X_t is a stationary process. We will see what transforms of the process Y_t lead to a stationary process, corresponding to various situations of the behavior of μ_t.

Example   Assume μ_t at two consecutive times are (always) almost

constant, that is,
$$\mu_t \approx \mu_{t-1}$$

Use $\beta_{0,t}$ to estimate μ_t by minimizing

$$\sum_t \left\{ [Y_{t-1} - \beta_{0,t}]^2 + [Y_t - \beta_{0,t}]^2 \right\}.$$

The LS estimator is

$$\beta_{0,t} = \frac{Y_t + Y_{t-1}}{2}.$$

(There is no theoretical reason to use Y_{t-1} instead of Y_{t+1}. But it makes slightly more sense to use the "past" rather than the "future". And this preference won't make systematic, practical differences.)

Then the detrended process is

$$Y_t - \beta_{0,t} = \frac{1}{2} \nabla Y_t.$$

Example   Do not assume $\mu_t \approx \mu_{t-1}$, but rather the "change" of μ_t is locally constant, that is,

$$\mu_{t+1} - \mu_t \approx \mu_t - \mu_{t-1}.$$

Estimate a mean series $\beta_{0,t}$ and a change-of-mean (or slope) series $\beta_{1,t}$ by minimizing

$$\sum_t \left\{ \left[Y_{t-1} - (\beta_{0,t} - \beta_{1,t}) \right]^2 + [Y_t - \beta_{0,t}]^2 + \left[Y_{t+1} - (\beta_{0,t} + \beta_{1,t}) \right]^2 \right\}.$$

The LS solution is

$$\beta_{0,t} = \frac{Y_{t-1} + Y_t + Y_{t+1}}{3}.$$

Then the detrended process is

$$Y_t - \beta_{0,t} = -\frac{1}{3} \nabla^2 Y_{t+1}.$$

Example   Assume μ_t is a RW, that is

$$\mu_t = \mu_{t-1} + \epsilon_t,$$

where ϵ is a white noise process independent of X. Then

$$\nabla Y_t = \nabla \mu_t + \nabla X_t = \epsilon_t + \nabla X_t.$$

Since X_t is stationary, ∇X_t is stationary. (Exercise: how can you prove this?) Then ∇Y_t is stationary.

Example   Instead of assuming μ_t is a RW, assume the change of μ_t is a RW. That is,

$$\nabla \mu_t = W_t = W_{t-1} + \epsilon_t.$$

Then,
$$\nabla Y_t = \nabla \mu_t + \nabla X_t = W_t + \nabla X_t.$$

Note W_t, a RW, is not stationary. Take difference again,

$$\nabla^2 Y_t = \nabla W_t + \nabla^2 X_t = \epsilon_t + \nabla^2 X_t$$

is stationary.

5 ARIMA models

ARIMA(p, d, q): $W_t = \nabla^d Y_t$ is ARMA(p, q).

Usually d is 1 or 2.

There are two layers of complexity here:
(1) d: how many steps of differences are needed?
(2) p, q: how complex (although stationary) is the resultant, difference process?

Since Y_t is nonstationary, it is pointless to study its infinite history (because its behavior changes with time). We'll assume $Y_{-m} = 0$ at time $t = -m < 1$, and study Y since that time.

Then, for example, if $d = 1$,

$$\nabla Y_t = W_t$$

we have

$$Y_t = Y_{t-1}+W_t = Y_{t-2}+W_{t-1}+W_t = \cdots = Y_{-m}+W_{-m}+W_{-m+1}+\cdots+W_t = W_{-m}+W_{-m+1}+\cdots+W_t$$

5.1 IMA(1,1)

If
$$\nabla Y_t = W_t = e_t - \theta e_{t-1},$$

then using $Y_t = \sum_{j=-m}^{t} W_j$ we can easily represent Y_t as a sum of e's. Then it's easy to work on the mean and variance, covariance functions.

Note   The example on page 93 emphasizes the non-stationarity of the Y_t process in several aspects.
(1) The derived variance and correlation functions depend on t;
(2) The autocorrelation stays high at large lags;
(3) The coefs of past e's do not die out.

Note   (5.2.6) looks like a MA process. What's the difference?

In A MA process, nonzero coefs extend back a fixed number of time steps; or they die out quickly so that effectively nonzero coefs extend back a fixed number of time steps.

In (5.2.6), significant coefs always extend back to a fixed starting point in time, i.e. Y_{-m}. This makes Y_t receive increasing amount of historical influence as t increases.

5.2   ARI(1,1)

$$\nabla Y_t = W_t = \phi W_{t-1} + e_t$$

We can not easily represent Y_t as a sum of e's as we did for IMA(1,1). Why can't we take the same approach?

Solution: (5.2.13), page 96.

5.3   Constant terms in ARIMA models

For ARIMA(p, d, q),

$$\nabla^d Y_t = W_t$$

is a stationary ARMA(p, q) process. Let's write it as X_t. We have assumed $E(X_t) = 0$. What if the mean of X_t is nonzero? Say

$$W_t = \mu + X_t$$

Two approaches:

(a) Study $U_t = W_t - \mu = X_t$, a zero-mean process.

(b) Introduce a constant term into the model:

$$W_t = \theta_0 + X_t$$

Note if W_t is written as a MA process, then θ_0 is μ; but if W_t is written with AR terms, then θ_0 is not μ. See page 97.

In general, if $E(W_t) = E(\nabla^d Y_t) \neq 0$, then

- if $d = 1$, this means Y_t has approximately constant increments, hence it has a mean that is a linear function of time t;

- if $d > 1$, by recursive thinking, $E(Y_t)$ is a d-th degree polynomial of time t.

6 Model specification for nonstationary processes

Informal signs on nonstationarity:

- Obvious trend in the Y series.

- Empirical ACF does not die out quickly enough: the pace of decrease is not nearly exponential, but more linear. (Note: for a nonstationary process, its ACF is not really defined. Nevertheless, the empirical ACF computed as if the process were stationary has this indicative use.)

Issues with nonstationarity:

(a) Transformation may help. Try log, square root, square, etc.

(b) Differencing, once or twice, often makes it much more stationary.

(c) Over-differencing is harmful!

(d) Therefore, carefully examine ∇Y_t; if it is really not enough, carefully examine $\nabla^2 Y_t$; and so one.

6.1 Pitfalls with differencing

Differencing appears to be the dominant idea in this course for dealing with nonstationarity. However, it should be noted that it can not address all nonstationarities. We will show this below.

Our goal is to transform Y_t to a stationary ARMA process X_t. Suppose after such a successful transformation, the model for X_t is

$$\phi(B)X_t = \theta(B)e_t$$

If this transformation is differencing, that is, $X_t = (1 - B)Y_t$, then the model for Y_t is actually

$$\phi(B)(1 - B)Y_t = \theta(B)e_t$$

This shows that if we have tried to model Y_t directly by ARMA, then the AR characteristic polynomial will have a unit root (root 1). In other words, existence of unit root in AR suggests we should do differencing.

Similarly, if differencing d times successfully transforms Y_t to a stationary X_t process, then the model for Y_t is

$$\phi(B)(1 - B)^d Y_t = \theta(B)e_t$$

that is, the AR model for Y_t has d unit roots.

From this observation we have this idea: test for unit root in an AR model (or the AR part of an ARMA model); if test says YES, then difference. Model the differenced series, test again; if test says YES, difference again.

Now, suppose Y_t is stationary (either the raw process or after an appropriate number of differencing), that is,

$$\phi(B)Y_t = \theta(B)e_t$$

What will happen if we difference this stationary Y_t (once more, if it is already a differenced series)?

Let $X_t = (1 - B)Y_t$. Notice

$$\phi(B)(1 - B)Y_t = \theta(B)(1 - B)e_t$$

The model for X_t is

$$\phi(B)X_t = \theta(B)(1 - B)e_t$$

We see this model (of X_t, i.e. differenced Y_t) has a unit root in its MA part! In other words, <u>existence of unit root in MA suggests we have over-differenced!</u>

We know a stationary process remains stationary after differencing. But the above shows the harm of over-differencing: it increases the MA order by one and, worse, introduces a unit root (which makes the MA model non-invertible).

This suggests we need to develop methods to test for unit root in a MA model as well.

What about nonstationarity due to a trend?

Suppose the true data-generating process is described by

$$Y_t = a_0 + a_1 t + X_t$$

where $X_t = \frac{\theta(B)}{\phi(B)}e_t$ is stationary. If we fit a linear trend and take it off from Y_t, then what's left is a stationary ARMA process.

If we instead take a difference, then

$$\nabla Y_t = a_1 + \frac{\theta(B)(1 - B)}{\phi(B)}e_t$$

we see the trend with time is knocked out by the differencing, but the model for the differenced series ∇Y_t is not as nice—its MA order is increased by 1 and it has a unit root.

If the correct model is

$$Y_t = a_0 + a_1 t + a_2 t^2 + X_t$$

where $X_t = \frac{\theta(B)}{\phi(B)} e_t$ is stationary, then

$$\nabla Y_t = a_1 + 2a_2 t + \frac{\theta(B)(1-B)}{\phi(B)} e_t$$

$$\nabla^2 Y_t = 2a_2 + \frac{\theta(B)(1-B)^2}{\phi(B)} e_t$$

The trend with time is removed by the two differencings, but the MA order of the model for the transformation is increased by 2 and the MA characteristic polynomial has 2 unit roots!

If a process becomes stationary after subtracting a trend, it's called a trend-stationary process. If differencing is required, it's called a difference-stationary process.

The preceding example shows that test for moving average unit root can help distinguishing trend-stationary and difference-stationary processes.

In this section, we assume the need to fit a trend function is somehow obvious, and it has been taken care of. In the following we focus on differencing.

To determine whether differencing is needed, and the correct number of differencing (i.e. the order of integration), we illustrate some ideas of unit root tests. The discussion is not complete nor systematic. This topic is not yet settled.

6.2   Unit roots in AR

For a nonstationary time series with a slowly decaying sample ACF and values near 1 at small lags, we use differencing to transform it into one with a rapidly decreasing sample ACF. The degree of differencing is largely determined by applying the difference operator repeatedly until the sample ACF of $\nabla^d Y_t$ decays quickly.

Since over-differencing is harmful, we need to be careful. Ideally, difference only when the AR model has unit roots.

Suppose Y_t are observations from the AR(1) model

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + e_t$$

We want to test H_0: $\phi_1 = 1$ vs H_1: $\phi_1 < 1$. To construct a test for H_0, write the model as

$$\nabla Y_t = Y_t - Y_{t-1} = \phi_0^* + \phi_1^* Y_{t-1} + e_t$$

where $\phi_0^* = \mu(1 - \phi_1)$ and $\phi_1^* = \phi_1 - 1$.

Find the ordinary least squares (OLS) estimator of ϕ_1^* by regressing ∇Y_t on 1 and Y_{t-1}. All standard routines will provide a T ratio which is the estimated ϕ_1^* divided by its standard error.

We use this T as test statistic, but it does not have a t distribution. Dickey and Fuller (some 30 years ago) derived the limit distribution as $n \to \infty$ under the unit root assumption $\phi_1^* = 0$. The .01, .05, and .10 quantiles of the limit distribution are -3.43, -2.86, and -2.57, respectively. Therefore, if $T = -3$, say, we can reject the unit root assumption at the 5% level, concluding $\phi_1^* < 0$, i.e. $\phi_1 < 1$, hence no differencing is needed.

This procedure can be extended to the case where $\{Y_t\}$ follows the $\mathrm{AR}(p)$ model with mean μ given by

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + e_t$$

This is re-written as

$$\nabla Y_t = \phi_0^* + \phi_1^* Y_{t-1} + \phi_2^* \nabla Y_{t-1} + \cdots + \phi_p^* \nabla Y_{t-p+1} + e_t$$

where $\phi_0^* = \mu(1 - \phi_1 - \cdots - \phi_p)$, $\phi_1^* = \sum_{i=1}^{p} \phi_i - 1$ and $\phi_j^* = -\sum_{i=j}^{p} \phi_i$, $j = 2, \ldots, p$. Testing the hypothesis of a unit root of the AR polynomial is equivalent to testing $\phi_1^* = 0$. As in the $\mathrm{AR}(1)$ example, ϕ_1^* can be estimated as the coefficient of Y_{t-1} in the OLS regression of ∇Y_t onto 1, Y_{t-1}, ∇Y_{t-1},..., ∇Y_{t-p+1}. The T ratio for the estimate of ϕ_1^* has exactly the same limit distribution as the one in the $\mathrm{AR}(1)$ case, therefore the same set of cutoff values can be used for the test.

6.3 Unit roots in MA

We discuss unit root test for a $\mathrm{MA}(1)$ model. More general cases are much more complicated.

Let $\{Y_t\}$ be observations from the $\mathrm{MA}(1)$ model

$$Y_t = e_t - \theta e_{t-1}$$

Davis and Dunsmuir (1996) showed that a test of $H_0 : \theta = 1$ vs $H_1 : \theta < 1$ can be performed to reject H_0 when

$$\hat{\theta} < -c_\alpha / n.$$

Values of c_α for typical α are given in Table 3.2 of Davis, Chen, and Dunsmuir (1995) $c_{.01} = 11.93$, $c_{.05} = 6.80$, $c_{.10} = 4.90$. (Cited in Brockwell and Davis 1996, p. 196. However, I couldn't find this reference.) For example, if $n = 50$, then the null hypothesis is rejected at level .05 if $\hat{\theta} < 1 - 6.80/50 = .864$.

7 Multiplicative seasonal ARMA models (SARMA)

Seasonaility is very common in time series data of economical, societal, and natural phenomena. Examples are monthly air traffic volume, quarterly unemployment rate, daily temperature. "Seasonality" here basically is synonymous with "periodicity".

Usually, "seasonality" is one type of "nonstationarity". For example, expected daily temperature certainly is nonstationary with a year. The nonstationarity here does not appear as a one-direction global "trend"; instead it is periodic. There are two ways to deal with this nonstationarity: fitting a deterministic, sinusoidal trend function, or differencing seasonally. It is not totally clear to me when we should use a periodic trend function and when we should use seasonal differencing. It might be the case that when some kind of "drift" exits in the process (e.g. with societal phenomenon), seasonal differencing should be used. In other words, periodic trend functions are used when, by reasonable arguments, the pattern just repeats over and over without much possibility to change in the long run.

After removing apparent seasonal nonstationarity, the residual often still show some correlation at the seasonal lags. We'll discuss how to model such seasonal correlations first, and come to differencing after that.

Denote the seasonal period by s. For example, for monthly sales of something, $s = 12$ is for annual cycles; for quarterly unemployment rate, $s = 4$ is for annual cycles.

Recall that in non-seasonal time series, we plot sample ACF and check whether it dampens with time quickly enough. If its rate of decrease is slow, that's indication of nonstationarity. Otherwise, we consider it a stationary process and model it with an ARMA formulation. With seasonal time series, the idea is the same. If the ACF at seasonal lags $s, 2s, 3s, \ldots$ does not decrease quickly enough (which is usually the case),

we need to remove the nonstationarity first. Now suppose that has been done, and ACF at seasonal lags shows some ARMA pattern.

7.1 MA

MA(2)
$$Y_t = \theta(B)\, e_t = (1 - \theta_1 B - \theta_2 B^2)\, e_t$$

where $\theta(B)$ satisfies the invertibility conditions, has $\rho(k)$ nonzero for $k = 0, 1, 2$ and zero for $k > 2$.

Similarly, make a seasonal MA model (taking $s = 12$) MA(2)$_{12}$

$$Y_t = \Theta(B)_{12}\, e_t = (1 - \Theta_1 B^{12} - \Theta_2 B^{24})\, e_t$$

where $\Theta(B)_{12}$ satisfies the invertibility conditions, has $\rho(k)$ nonzero for $k = 0, 12, 24$ and zero otherwise (speaking of non-negative lag k only).

It is perhaps always the case that correlation exists not only at the seasonal lags, but between neighboring times steps as well. So we could do this:

$$Y_t = e_t - \theta e_{t-1} - \Theta e_{t-12}.$$

This has non-zero ρ at lags 0, 1, 11, 12. This is a MA(12) with many θs set to 0.

A widely used formulation constructs the model by "multiplication". For example a MA(1) $\times$ MA(1)$_{12}$ is defined by

$$Y_t = (1 - \theta B)(1 - \Theta B^{12})\, e_t = (1 - \theta B - \Theta B^{12} + \theta\Theta B^{13})\, e_t$$

This model has nonzero ρ at lags 0, 1, 11, 12, 13. This is a MA(13) model with many θs set to 0.

This <u>multiplicative</u> formulation has bee very popular. It has proven to be adequate for many application (although not all seasonality can be modeled this way). Advantages of the multiplicative formulation:

1. It's <u>parsimonious</u>. For example, the MA(1) $\times$ MA(1)$_{12}$ model above uses only 2 parameters to formulate a MA(13), yet capturing both seasonal and nonseasonal correlations.

2. The characteristic polynomial is easier to discuss (because it is already factorized) than a general MA(13).

3. It is easily extended, for example

$$Y_t = (1 - \theta_1 B - \theta_2 B)(1 - \Theta_1 B^{12} - \Theta_2 B^{24}) e_t$$

uses 4 parameters to construct a MA(26).

Exercise   At what lags does this model have nonzero correlation?

7.2   AR

AR(1)

$$\phi(B)Y_t = (1 - \phi B)Y_t = e_t$$

where $\phi(B)$ satisfies the causality conditions, has $\rho(k)$ nonzero at $0, 1, 2, 3, \ldots$.

Similarly, make a seasonal AR$(1)_{12}$

$$\Phi(B)_{12}Y_t = (1 - \Phi B^{12})Y_t = e_t$$

where $\Phi(B)$ satisfies the causality conditions, has nonzero correlations $\rho(k)$, $k = 0, 12, 24, 36, 48, \ldots$.

Make a simple multiplicative AR mode, AR$(1) \times$ AR$(1)_{12}$,

$$(1 - \phi B)(1 - \Phi B^{12})Y_t = e_t$$

This model as nonzero correlation at all lags, due to the AR(1) component.

7.3   ARMA

In general, we can define an ARMA$(p, q) \times (P, Q)_s$ as

$$\phi(B)\Phi(B)Y_t = \theta(B)\Theta(B)e_t$$

that is,

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - \Phi_1 B^s - \cdots - \Phi_P B^{Ps})Y_t = (1 - \theta_1 B - \cdots - \theta_q B^q)(1 - \Theta_1 B^s - \cdots - \Theta_Q B^{Qs})Y_t$$

This is a special ARMA$(Ps + p, Qs + q)$ model with $p + P + q + Q$ (rather than $p + Ps + q + Qs$) parameters.

Note: writing $\phi(B)\Phi(B)$ is the same as $\Phi(B)\phi(B)$. Similarly for the MA part.

Seasonal ARMA models is also called SARMA.

Example   Exhibit 10.3, page 231.

Example   Exhibit 10.4, page 232.

8 Integrated seasonal ARMA models (SARIMA)

Suppose the process is stationary after D seasonal differencing and d nonseasonal differencing, then we model the differenced form, $(1-B)^d(1-B^s)^D Y_t$, by a SARIMA model:

$$\phi(B)\Phi(B)\nabla^d\nabla_s^D Y_t = \theta(B)\Theta(B)\,e_t$$

where $\nabla^d = (1-B)^d$, $\nabla_s^D = (1-B^s)^D$. This model is denoted by $\mathrm{ARMA}(p,d,q) \times (P,D,Q)_s$.

Note
1. Usually d and D are at most 2.

2. Since $\nabla\nabla_s Y_t$ is the same as $\nabla_s\nabla Y_t$, it's easy to see that $\nabla^d\nabla_s^D Y_t$ is the same as $\nabla_s^D\nabla^d Y_t$.

Example   Section 10.4, page 234.

This example nicely shows various considerations in the whole procedure of model specification, fitting, and checking. Some of the points are summarized below.

1. Always, plot the series and look at it.

2. If the whole series shows a global trend, do a regular (i.e., $(1-B)$) difference. Plot the differenced series.

3. After removing global trends by non-seasonal difference, check and see if there is seasonality. If there is, do a seasonal difference, then plot the series and see if seasonality is still clear and strong. If it is, difference again, and check again.

4. Always, compute and plot sample ACF along with any examination of the series. If <u>significant</u> (i.e. statistically non-zero) r (either at lags $1, 2, \ldots$ or lags $1s, 2s, \ldots$) does not die out quickly enough, it's indication of non-stationarity.

5. Once the series looks reasonably stationary, tentatively determine orders p, q, P, Q. Fit the model.

6. What to look at for diagnostics?

   (a) Plot the residuals. Watch for any systematic patterns.

   (b) Plot the sample ACF of the residuals.
   This is a further check of whether there is any systematic pattern in the residuals that can be described by a model. If the model is adequate, ACF of the residuals should be insignificant (at all lags).

If there are a few significant ACF, consider the following:

   i. How significant? Very much so or just marginally so?

  ii. Can you think of a reason/explanation for this significant ACF?

 iii. How many lags are significant? If there are one or two out of a whole lot, there significance could have happend by pure chance.

If the significance (or lack thereof) of the residual ACF is not obvious, perform a Ljung-Box test.

(c) Check normality of the residuals.

   i. Histogram of residuals.

  ii. Normality (QQ) plot.

 iii. Formal test of normality.

(d) Consider overfitting the model by increasing the order $(p,\ q,\ P,\ Q)$ by 1. Compare the re-fitted model with the original model, paying attention to

   i. Have the common coefficients changed a lot?

  ii. Is the additional coefficient significant?

 iii. Has the $\hat{\sigma}_e^2$ changed a lot?

 iv. Has the log-likelihood achieved by the estimates changed (improved) a lot?

  v. How has AIC changed?